EUROPEAN ASSOCIATION FOR MACHINE TRANSLATION

EAMT2018

Proceedings of the

21st Annual Conference of the European Association for Machine Translation

28–30 May 2018
Universitat d'Alacant
Alacant, Spain

Edited by
Juan Antonio Pérez-Ortiz
Felipe Sánchez-Martínez
Miquel Esplà-Gomis
Maja Popović
Celia Rico
André Martins
Joachim Van den Bogaert
Mikel L. Forcada

Organised by

Universitat d'Alacant
Universidad de Alicante

transducens
research group

ISBN: 978-84-09-01901-4

Gaussian Processes for Text Regression

Daniel Beck[1]
School of Computing and Information Systems
The University of Melbourne, Australia
d.beck@unimelb.edu.au

This thesis deals with the general problem of predicting numerical indicators from textual data. This task, which we call Text Regression, arises in a range of different applications in Natural Language Processing (NLP). For instance, in Quality Estimation (QE) (Blatz et al., 2004; Specia et al., 2009), sentences generated from Machine Translation (MT) systems are evaluated according to a task-based metric such as post-editing effort or time. In Emotion Analysis (EA) (Strapparava and Mihalcea, 2007), natural language sentences are assigned with numerical scores mapping the strength of a particular emotion (or a set of emotions).

Standard approaches for Text Regression rely on architectures similar to the ones used in classification tasks. These use engineered features and/or simple text representations such as bag-of-words (BOW), and make predictions in the form of single point estimates. These simplifying assumptions ignore important aspects of the data. Representations such as BOW ignore structural aspects of sentences and fails to capture structural linguistic phenomena such as word order. Point estimate predictions lack uncertainty information on the predicted variable, which can help subsequent decision making and is particularly important when annotations are noisy (such as post-editing time in QE).

The goal of this thesis is to advance the state-of-the-art in Text Regression by improving these two aspects: improved text representations and better uncertainty modelling in the response variables. In order to achieve that goal we propose to use Gaussian Processes (GPs) (Rasmussen and Williams, 2006) as the regression model. GPs are a Bayesian kernelised framework which is considered the state-of-the-art in regression (Hensman et al., 2013). Perhaps surprisingly, GPs were not widely investigated in the context of NLP applications.[2] Therefore a secondary goal of this thesis is to disseminate GPs in the NLP community, in particular for regression tasks.

The theory behind Gaussian Processes regression makes it ideal to solve the two problems mentioned above. Since it models response variables as well-calibrated distributions, it naturally provides a measure of uncertainty over the predictions. Furthermore, by employing kernels as the underlying learning component, we can incorporate complex text representations through what we named *structural kernels*. Combining with the efficient model selection procedures provided by GPs, we show in this thesis how to essentially learn representations by enabling richer kernel parameterisations. In this thesis, we focus on string kernels (Lodhi et al., 2002; Cancedda et al., 2003) and tree kernels (Collins and Duffy, 2001; Moschitti, 2006) but the theory can easily be extended to other kinds of structures such as graph kernels (Vishwanathan et al., 2010).

We benchmark our approach in two Text Regression applications. The first one is Emotion Analysis, where we use a GP model with a soft string kernel using word embeddings for similarity calculation between words. We show that this proposed model can obtain better results compared to simpler baselines. For this task, we also propose a multi-task model which leverages multiple emotional labels and show how we can inspect GP

[1]This thesis was written while the author was a Ph.D. student at The University of Sheffield, United Kingdom.

[2]Notable exceptions are Polajnar et al. (2011) and Cohn and Specia (2013).

Pérez-Ortiz, Sánchez-Martínez, Esplà-Gomis, Popović, Rico, Martins, Van den Bogaert, Forcada (eds.)
Proceedings of the 21st Annual Conference of the European Association for Machine Translation, p. 5–6
Alacant, Spain, May 2018.

hyperparameters to cluster similar emotions.

The second benchmark is Machine Translation Quality Estimation. In this task, we show that can obtain better results compared to baselines while also providing uncertainty estimates for predictions. More important, we show how to employ the predictive distributions in an asymmetric risk scenario, where over and underestimates of post-editing time have different costs. This is an example application where propagating full uncertainty information can be beneficial for further decision making in a translation pipeline. As another application example, we also show how to use uncertainty estimates to annotate QE datasets via active learning.

Finally, as mentioned before, this thesis also has the goal of disseminating Gaussian Processes among the NLP community. By providing the theoretical grounds and showcasing its application in two benchmarks, we hope that it will serve as a starting point for other NLP problems in the future.

Access to the full thesis is open and available at the White Rose eTheses repository (`etheses.whiterose.ac.uk/17619`).

Acknowledgements

Daniel Beck was supported by funding from CNPq/Brazil (No. 237999/2012-9). The author would also like to thank his Ph.D. supervisors, Lucia Specia and Trevor Cohn, and his thesis examiners, Andreas Vlachos and Xavier Carreras.

References

Blatz, John, Erin Fitzgerald, George Foster, Simona Gandrabur, Cyril Goutte, Alex Kulesza, Alberto Sanchis, and Nicola Ueffing. 2004. Confidence estimation for machine translation. In *Proceedings of the 20th Conference on Computational Linguistics*, pages 315–321.

Cancedda, Nicola, Eric Gaussier, Cyril Goutte, and Jean-Michel Renders. 2003. Word-Sequence Kernels. *The Journal of Machine Learning Research*, 3:1059–1082.

Cohn, Trevor and Lucia Specia. 2013. Modelling Annotator Bias with Multi-task Gaussian Processes: An Application to Machine Translation Quality Estimation. In *Proceedings of ACL*, pages 32–42.

Collins, Michael and Nigel Duffy. 2001. Convolution Kernels for Natural Language. In *Proceedings of NIPS*, pages 625–632.

Hensman, James, Nicolò Fusi, and Neil D. Lawrence. 2013. Gaussian Processes for Big Data. In *Proceedings of UAI*, pages 282–290.

Lodhi, Huma, Craig Saunders, John Shawe-Taylor, Nello Cristianini, and Chris Watkins. 2002. Text Classification using String Kernels. *The Journal of Machine Learning Research*, 2:419–444.

Moschitti, Alessandro. 2006. Making Tree Kernels practical for Natural Language Learning. In *EACL*, pages 113–120.

Polajnar, Tamara, Simon Rogers, and Mark Girolami. 2011. Protein interaction detection in sentences via Gaussian Processes: a preliminary evaluation. *International Journal of Data Mining and Bioinformatics*, 5(1):52–72, jan.

Rasmussen, Carl Edward and Christopher K. I. Williams. 2006. *Gaussian processes for machine learning*, volume 1. MIT Press Cambridge.

Specia, Lucia, Nicola Cancedda, Marc Dymetman, Marco Turchi, and Nello Cristianini. 2009. Estimating the sentence-level quality of machine translation systems. In *Proceedings of EAMT*, pages 28–35.

Strapparava, Carlo and Rada Mihalcea. 2007. SemEval-2007 Task 14 : Affective Text. In *Proceedings of SemEval*, pages 70–74.

Vishwanathan, S. V. N., Nicol N. Schraudolph, Risi Kondor, and Karsten M. Borgwardt. 2010. Graph Kernels. *Journal of Machine Learning Research*, 11:1201–1242.